

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 5, 2001, 15:23:33 ; Search time 58.44 Seconds
(Without alignments)
215.709 Million cell updates/sec

Title: US-09-593-316-4
Perfect score: 2008
Sequence: 1 MNKGVKLVLSMLVSTIVV.....IKLVKMSQTKRYNVRNV 368

Scoring table:
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2008	100.0	368	1	GATR_BOVIN
2	1717.5	85.5	371	1	GATR_BOVIN
3	1476.5	73.5	394	1	GATR_MOUSE
4	655	32.6	354	1	BCAT_HUMAN
5	115.5	5.8	505	1	AGAR_ALTRAT
6	102.5	5.1	559	1	PAGT_BOVIN
7	101.5	5.1	559	1	PAGT_HUMAN
8	97.5	4.9	446	1	GLNA_METP
9	95	4.7	1057	1	POL_STVAT
10	94.5	4.7	559	1	PAGT_RAT
11	94.5	4.7	2410	1	POL1_BAYM
12	93	4.6	408	1	YZ33_METJA
13	92.5	4.6	1418	1	CAI2_HUMAN
14	90	4.5	522	1	HEX1_ENTHI
15	90	4.5	708	1	TALA_SV40
16	89.5	4.5	478	1	MANC_ECOLI
17	89.5	4.5	2054	1	YCP2_PINTH
18	89	4.4	688	1	TALA_POVIC
19	89	4.4	1826	1	SUTS_HUMAN
20	88.5	4.4	369	1	CAI2_CHICK
21	88.5	4.4	1459	1	CAI2_MOUSE
22	88	4.4	567	1	YB28_METJA
23	87	4.3	754	1	YVY1_CABEI
24	86	4.3	355	1	YVY1_METJA
25	86	4.3	1056	1	POL_BIV06
26	86	4.3	1056	1	POL_BIV27
27	85.5	4.3	385	1	YGLW_YEAST
28	85.5	4.3	446	1	GLNA_METV
29	85	4.2	541	1	STR_HELPY
30	84	4.2	454	1	GLNA_METJA
31	84	4.2	1018	1	YCI4_METJA
32	84	4.2	1826	1	SUTS_RABIT
33	83.5	4.2	332	1	METE_AERPE

ALIGNMENTS

RESULT 1	ID	GATR_BOVIN	STANDARD:	PRT:	368 AA.
AC	P14769.				
DT	01-APR-1990 (Rel. 14, Created)				
DT	01-APR-1990 (Rel. 14, Last sequence update)				
DT	30-MAY-2000 (Rel. 39, Last annotation update)				
DE	N-ACETYLACTOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE (EC 2.4.1.151)				
DE	(GALACTOSYLTRANSFERASE) (UDP-GALACTOSE:BETA-D-GALACTOSYL-1,4-N-ACETYL-				
DE	D-GLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE).				
GN	GTAL1.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Eumetazomia;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
OX	NCBI_TaxID:9913;				
RP	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE:89340543; PubMed:2503516;				
RA	Torjasse D.H., Shaper J.H., van den Fijnden D.H., van Tunen A.J.,				
RA	Shaper N.L.;				
RT	*Bovine alpha 1-->3-galactosyltransferase: isolation and				
RT	characterization of a cDNA clone. Identification of homologous				
RT	sequences in human genomic DNA.*;				
RL	J. Biol. Chem. 264:14290-14297(1989).				
CC	-1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN				
CC	ACCEPTOR MOLECULE (R).				
CC	-1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-				
CC	N-ACETYL-D-GLUCOSAMINYL-R = UDP + ALPHA-D-GALACTOSYL-1,3-				
CC	BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.				
CC	-1- COFACTOR: MANGANESE.				
CC	-1- PATHWAY: GLYCOSYLATION.				
CC	-1- SUBCELLULAR LOCATION: TYPE 1) MEMBRANE PROTEIN. MEMBRANE-BOUND				
CC	FORM IN TRANS CISTERNAE OF GOLGI.				
CC	-1- DISEASE: AUTOIMMUNE DISEASE (ANTIBODIES AGAINST ENZYMATIC				
CC	PRODUCT).				
CC	-1- SIMILARITY: STRONG TO BLOOD GROUP AB TRANSFERASE.				
CC	-1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN				
CC	GLYCOSYLTRANSFERASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.jsb.slb.ch/announce/				
CC	or send an email to license@jsb.slb.ch).				
DR	EMBL: J04989; AAA30558.1; ..				
DR	PIR: A44785; A44785.				
KM	Signal-anchor; Glycosyltransferase; Glycoprotein; Transmembrane;				
KM	Signal-anchor; Golgi stack; Manganese.				
FT	DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).				
FT	DOMAIN 7 22 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).				
FT	DOMAIN 23 368 LUMENAL, CATALYTIC (POTENTIAL).				
FT	CARBOHD 293 293 N-LINKED (GLCNAC:...) (POTENTIAL).				

SEQUENCE 468 AA: 43246 MW: 58650066737300033 CRC64:

Query Match 100.0%: Score 2008; ID 1; Length 368;

Host Local Similarity 100.0%: Pred. No. 1,9e-154;

Matches 468: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MNVKGAVLLSMVSVIVAVWVWYHSPGSLFWINPSPNPGVGSSTQKGMWLPWPN 60
 1 MNVKGAVLLSMVSVIVAVWVWYHSPGSLFWINPSPNPGVGSSTQKGMWLPWPN 60
 1 MNVKGAVLLSMVSVIVAVWVWYHSPGSLFWINPSPNPGVGSSTQKGMWLPWPN 60
 61 GYHEEDGIDNEEKQNDNEKQNDNEKQNDNEKQNDNEKQNDNEKQNDNEKQND 120
 61 GYHEEDGIDNEEKQNDNEKQNDNEKQNDNEKQNDNEKQNDNEKQNDNEKQND 120
 121 YVAKOKITVGLTVAVAGRYTEHLEETLSANSHFWGVHVFYIMVDVSRMPLLE 180
 121 YVAKOKITVGLTVAVAGRYTEHLEETLSANSHFWGVHVFYIMVDVSRMPLLE 180
 121 YVAKOKITVGLTVAVAGRYTEHLEETLSANSHFWGVHVFYIMVDVSRMPLLE 180
 181 LRSKRVKIKRKRWDQISMRKKTIGEHVAHLOHEVDFEPCMDVDFVQKFGVETIG 240
 181 LRSKRVKIKRKRWDQISMRKKTIGEHVAHLOHEVDFEPCMDVDFVQKFGVETIG 240
 241 ESYVAGVAMWYKADPNDFYERKESAAVIFPGEGDFYTHAIFGGTFYVNLTF 300
 241 ESYVAGVAMWYKADPNDFYERKESAAVIFPGEGDFYTHAIFGGTFYVNLTF 300
 401 GILKOKKNDIEAOMHDESHLNKFTLLNKPTKLLSPYGMVYHGLVADIKLVKMS 360
 401 GILKOKKNDIEAOMHDESHLNKFTLLNKPTKLLSPYGMVYHGLVADIKLVKMS 360
 461 YVAVKNNV 468
 461 YVAVKNNV 468

RESULT 2
 GATK_P16 STANDARD: PRT: 471 AA.
 ID GATK_P16 STANDARD: PRT: 471 AA.
 AC P50127;
 DT 01-OCT-1996 (rel. 44, created)
 DT 01-OCT-1996 (rel. 44, last sequence update)
 DT 01-OCT-1996 (rel. 44, last annotation update)
 DE N-ACETYLTRANSFERASE (EC 2.4.1.151)
 DE (GALACTOSYLTRANSFERASE) (UDP-GALACTOSE: BETA-D-GALACTOSYL-1,4-N-ACETYL-
 DE D-GALACTOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE).
 GN GATK.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Cetartia; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.
 IX MBL_042104824;
 IX 11;
 IX STRAIN YORKSHIRE TISSUE Spleen;
 IX MEDLINE:95104914; PubMed:7528726;
 IX Strickland K.M., Gu P., Prosser A.F., Gustavsson L., Andersson L.,
 IX Gustavsson K.,
 IX "DNA sequence and chromosome localization of pig alpha 1,3
 IX galactosyltransferase".
 IX Immunogenetics 41:101-105(1995).
 IX 121
 IX SEQUENCE FROM N.A.
 IX TISSUE-Spleen, and Liver;
 IX Smith M.S., Jankowski P.L., Bennett M.M., Moutouris E.,
 IX McKnight J.F.C.,
 IX Submitted (Feb-1995) to the EMBL/Genbank/DBJ databases.
 IX FUNCTION: TRANSFER OF GALACTOSE FROM UDP GALACTOSE TO AN
 IX ACCEPTOR MOLECULE (R).
 IX CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-
 IX N-ACETYL D-GALACTOSAMINIDE -> UDP + ALPHA-D-GALACTOSYL-1,3-
 IX BETA-D-GALACTOSYL-1,4-N-ACETYL D-GALACTOSAMINIDE-1,3-
 IX D-GALACTOSAMINIDE.
 IX COFACTOR: MANGANESE.
 IX PATHWAY: GLYCOSYLATION.

1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 FORM IN TRANS GISTERN OF GOLGI.
 1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.
 1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
 GLYCOSYLTRANSFERASES.

This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation.
 The European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.ebi.ac.uk/submit/>
 or send an email to license@ebi.ac.uk).

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

DR FMBL: L36152; AAA58775.1; -

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 CC NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96046769; PubMed=2510162;
 RA Larsen R.D., Rajan V.P., Ruff M.M., Kukowska-Latallo J.,
 RA Cummings R.D., Lowe J.B.,
 RT Isolation of a cDNA encoding a murine
 RT UDPgalactose:beta-D-galactosyl-1,4-N-acetyl-D-glucosaminide
 RT alpha-1,3-galactosyltransferase: expression cloning by gene
 RT transfer.*;
 RT Proc. Natl. Acad. Sci. U.S.A. 86:8227-8231(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92184813; PubMed=1544928;
 RA Joziase D.H., Shaper N.L., Kim D., van den Bijnden D.H., Shaper J.H.,
 RT Murine alpha 1,3-galactosyltransferase. A single gene locus
 RT specifies four isoforms of the enzyme by alternative splicing.*;
 RT J. Biol. Chem. 267:5534-5541(1992).
 CC -1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN
 CC ACCEPTOR MOLECULE (R).
 CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-
 CC N-ACETYL-D-GLUCOSAMINYL-R + UDP + ALPHA-D-GALACTOSYL-1,3-
 CC BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
 CC -1- COFACTOR: MANGANESE.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 CC FORM IN TRANS CISTERNAE OF GOLGI.
 CC -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.
 CC -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
 CC GLYCOSYLTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M26925; AAA37657.1;
 DR EMBL: M85153; AAA37711.1;
 DR PIR: A34417; A34417.
 DR MCD: M61:95704; Gytal.
 KM Signal-anchor: Glycosyltransferase; Glycoprotein; Transmembrane;
 KM Signal-anchor: Golgi stack; Alternative splicing; Manganese.
 FT DOMAIN 1 41 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 42 60 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 61 394 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 83 83 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT CARBOHYD 319 419 N-LINKED (GLCNAc. . .) (POTENTIAL).
 FT VARSPLIC 1 35 MISSING (IN SHORTER ISOFORM).
 FT VARSPLIC 62 62 R -> SPDSFLMIYHETK (IN SHORTER ISOFORM).
 FT SEQUENCE 394 AA: 46475 MW: 7766831640DIBBF7 CRG64.
 Query Match 73.5%; Score 1476.5; DB 1: Length 394;
 Best local similarity 72.2%; Pred No. 1-2e-111;
 Matches 268: Conservative 41; Mismatches 47; Indels 15; Gaps 3;
 QY 1 MNVGRKVLISMLVSTVIVFWYJHSPEGLFWINPSNPEVSGSSITQKGMPEPRFNN 60
 DB 36 MNVGRKVLISMLVSTVIVFWYJHSPEGLFWINPSNPEVSGSSITQKGMPEPRFNN 60
 QY 61 GYH--EEGDINEERQRNEDE-SKLKLSDMNPRKREPVNTMKKAPVVEGYTNAV 117
 DB 84 GYH--EEGDINEERQRNEDE-SKLKLSDMNPRKREPVNTMKKAPVVEGYTNAV 117
 QY 118 LNNYAKOKITVGLTFAFGVYIEHYLEELTSANKHFWGHPVIFVYDVSHPPLIE 177
 DB 144 LKYYATOKLITGLTFAFGVYIEHYLEELTSANKHFWGHPVIFVYDVSHPPLIE 177

QY 178 LGPLRSKVFYFKIKRERKMODISMRKKTIGCEHYAHIOHFVDFLEPCMDVQVDPKRGVE 237
 DB 204 LNPJLSLOVFEIRSEKRWODISMRKKTIGCEHYAHIOHFVDFLEPCMDVQVDPKRGVE 263
 QY 238 TLGEVQQLQAMWYRKADPNPTFYERKRESAAVYIPEGCEDFYHANAIFGGTPTVLNITQ 297
 DB 264 TLGLVQQLQAMWYRKADPNPTFYERKRESAAVYIPEGCEDFYHANAIFGGTPTVLNITQ 323
 QY 298 CPGKILDKKNDIEAOWHDESHLNKYPFLNPKTKIISPPYCWYHIGLPAHILKYSWQ 357
 DB 324 CPGKILDKKNDIEAOWHDESHLNKYPFLNPKTKIISPPYCWYHIGLPAHILKYSWQ 383
 QY 358 TKENYVANNV 368
 DB 384 TKENYVANNV 394
 RESULT 4
 ID BGAT_HUMAN STANDARD: PRT: 354 AA.
 AC P16442;
 DT 01-AUG-1990 (Ref. 15, Created)
 DT 01-DEC-1992 (Ref. 24, Last sequence update)
 DT 01-OCT-2000 (Ref. 40, Last annotation update)
 DE HISTO-BLOOD GROUP ABO SYSTEM TRANSFERASE (NAGAT) [INCLUDES:
 DE GLUCOPROTEIN-FUCOSYLTRANSFERASE (FC 2.4.1.40) (FUCOSYLGLUCOPROTEIN
 DE ACETYLALACTOSAMINYLTRANSFERASE (FC 2.4.1.40) (FUCOSYLGLUCOPROTEIN
 DE ALPHA-N-ACETYLALACTOSAMINYLTRANSFERASE) (HISTO-BLOOD GROUP A
 DE TRANSFERASE) (A TRANSFERASE); GLYCOPROTEIN-FUCOSYLALACTOSIDE ALPHA-
 DE GALACTOSYLTRANSFERASE (FC 2.4.1.37) (FUCOSYLGLUCOPROTEIN 3 ALPHA-
 DE GALACTOSYLTRANSFERASE) (HISTO-BLOOD GROUP B TRANSFERASE) (B
 DE TRANSFERASE)].
 GN ABO.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 CC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=90110098; PubMed=2104828;
 RA Yamamoto F.-I., Marken J., Tsuji T., White T., Clausen H.,
 RA Hakomori S.-I.,
 RT Cloning and characterization of DNA complementary to human UDP-
 RT GalNAc-6-epi alpha 1-3-galactose 4-epimerase (histo-blood
 RT group A transferase) mRNA.*;
 RL J. Biol. Chem. 265:1146-1151(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90238543; PubMed=2333095;
 RA Yamamoto F.-I., Clausen H., White T., Marken J., Hakomori S.-I.,
 RT Molecular genetic basis of the histo-blood group ABO system.*;
 RL Nature 345:229-233(1990).
 RN [3]
 RP CHARACTERIZATION.
 RX MEDLINE=91035461; PubMed=2121736;
 RA Yamamoto F.-I., Hakomori S.-I.,
 RT Sugar-nucleotide donor specificity of histo-blood group A and B
 RT transferases is based on amino acid substitutions.*;
 RL J. Biol. Chem. 265:19257-19262(1990).
 CC -1- FUNCTION: THIS PROTEIN IS THE BASIS OF THE ABO BLOOD GROUP SYSTEM.
 CC THE HISTO-BLOOD GROUP ABO INVOLVES THREE CARBOHYDRATE ANTIGENS: A,
 CC B, AND H. A, B, AND AB INDIVIDUALS EXPRESS A GLYCOSYLTRANSFERASE
 CC ACTIVITY THAT CONVERTS THE H ANTIGEN TO THE A ANTIGEN (BY ADDITION
 CC OF UDP-GALNAc) OR TO THE B ANTIGEN (BY ADDITION OF UDP-GAL),
 CC WHEREAS O INDIVIDUALS LACK SUCH ACTIVITY.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + GLYCOPROTEIN
 CC ALPHA-L-FUCOSYL-(1,2)-D-GALACTOSE = UDP + N-ACETYL-D-GALACTOSE.
 CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + GLYCOPROTEIN ALPHA-L-FUCOSYL-
 CC (1,2)-D-GALACTOSE = UDP + GLYCOPROTEIN ALPHA-D-GALACTOSYL-(1,3)-
 CC [ALPHA-L-FUCOSYL-(1,2)]-D-GALACTOSE.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND

[illegible]

```

01 334 SPEYGMWYH-IGLIADH-KLAKMSWQTKFYVAVRN 366
02 111111111111111111111111111111111111
03 320 SPEYLMIDQILLGMPVAVLKIKLRTFAVPRKNIQAVRN 353
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
```

QY 225 DVDQVQ-----DKFVETLCESV-----AQLQAMRY--K 252
 DB 287 EDNKRKGRSLEATLMDIDYSLALGELTKPHVISEYGGRRRPMENAPMSALRDMWELKT 346
 QY 253 ADNDPFLYERKRSKA---AYIPGEGDFFYHAAIFG---GTPQVLTNTOCEFGILK 304
 DB 347 ASPLMLQFIISKPSVLSLTFVPI-----KALMKTADGTFPYMRLRLRQ-----K 392
 QY 305 DKRNDIEAQM 314
 DB 393 EADNETGENM 402

RESULT 6
 PACT_BOVIN
 ID PACT_BOVIN STANDARD: PRT: 559 AA.
 AC 007537;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 39, Last annotation update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE POLYPEPTIDE N-ACETYLGLACTOSAMINYLTRANSFERASE (EC 2.4.1.41) (PROTEIN-
 DE UDP ACETYLGLACTOSAMINYLTRANSFERASE) (UDP-GALNAC:POLYPEPTIDE, N-
 DE ACETYLGLACTOSAMINYLTRANSFERASE) (GALNAC-T1).
 GN GALNT1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN (1)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 41-74.
 RC TISSUE=Intestine, and Colostrum;
 RA MEDLINE=93286099; PubMed=7685345;
 RA Homa F.L., Hollander T., Lehman D.J., Thomsen D.R., Elhammer A.P.;
 RT Isolation and expression of a cDNA clone encoding a bovine UDP-
 RT GALNAC:polypeptide N-acetylglactosaminyltransferase.";
 RL J. Biol. Chem. 268:12609-12616(1993).
 RN (2)
 RP SEQUENCE OF 41-559 FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Placenta, and Colostrum;
 RX MEDLINE=93366815; PubMed=8360184;
 RX Hagen F.K., van Muijkhuysen B., Tabak L.A.;
 RA "Purification, cloning, and expression of a bovine UDP-GalNAC:
 RT polypeptide N-acetyl galactosaminyltransferase.";
 RL J. Biol. Chem. 268:18960-18965(1993).
 CC - FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
 CC OLIGOSACCHARIDE BIOSYNTHESIS. THE TRANSFER OF AN N-ACETYL-D-
 CC GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
 CC PROTEIN RECEPTOR. THE ENZYME MAY PREFERENTIALLY TRANSFER TO
 CC THREONINE.
 CC - CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =
 CC UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.
 CC - COFACTOR: MANGANESE AND CALCIUM.
 CC - PATHWAY: GLYCOSYLATION.
 CC - SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI.
 CC - TISSUE SPECIFICITY: COLOSTRUM CONTAINS A SOLUBLE FORM.
 CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: L07780: AAA30532.1: -;
 CC EMBL: L17437: AAA68489.1: -;
 CC Interpro: IPR000772: -;
 CC Interpro: IPR001173: -;
 CC Pfam: PF00535: Glycos.transf.2; 1.
 CC Pfam: PF00552: Richn_B lectin; 1.

KW Transferase: Glycosyltransferase; Transmembrane; Signal anchor;
 KM Goldi stack: Glycoprotein; Manganese; Calcium;
 FT PROPEP 1 40 REMOVED IN SOLUBLE POLYPEPTIDE
 FT CHAIN 41 559 N-ACETYLGLACTOSAMINYLTRANSFERASE.
 FT POLYPEPTIDE N-
 FT ACETYLGLACTOSAMINYLTRANSFERASE, SOLUBLE
 FT FORM.
 FT DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).
 FT TRANSFER 9 28 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 29 559 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 117 117 POTENTIAL.
 FT CARBOHYD 118 118 POTENTIAL.
 FT CARBOHYD 119 119 POTENTIAL.
 FT CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 144 288 POTENTIAL.
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 559 AA; 64192 MW; E3E538C4DE569840 CRC64;

Query Match 5.1%; Score 102.5; DB 1; Length 559;
 Best local Similarity 20.5%; Pred. No. 0.72;
 Matches 80; Conservative 52; Mismatches 131; Indels 127; Gaps 22;

QY 63 HEEDGIDNEFKRQNEDESKIKLSDFNPFK--RPVYVTRTKAKAPVWEGTYNRAVLIN 120
 DB 55 HEGPGEKRPVVIPIKDOEKMKEMKINQFNLMASEMIALNSLPVRLGCKTVYVUN 114
 QY 121 -----YAAKQITVCLTFEAV--GRYIEHYLEFLTSANKHHPVIFVIMQDV 170
 DB 115 LPTTSVIVFHNAMSTLRTVHSVINRSPRHLEI-----VLVIDA 157
 QY 171 SRMPLIELGLRNF-----KYKIKPKRWODISMRRK-----T 205
 DB 158 SERDFLK-RPLESVYKLVKVPVAVIMEQR-SGLIFARLKGAANSQGYITELDAHCECT 215
 QY 206 IG--EHIVAHIOHEVPLFCMDVDVFOQK-----GVE-TLGSVNOLOAMVKKADPNDF 258
 DB 216 VGLPEPLARIKHDKRTVCPITDIVISDTFELMAGSDMTYGSFNRKLNFRVPIVQRHM 275
 QY 259 TYRRRESAAVIFPG-----EGDFY-----YHAA--IFGTPVOVLTNTOCE-- 298
 DB 276 --DRKKGDRI-LVPRPTMAAGLFSIDROYFOELGTGYDGMQIMGENSELISPRIMQCG 332
 QY 299 -----FKG-----ILKQKNDIEQWDESHLNKVFLLNKP--TK 331
 DB 333 TLEIVTCSHGVFRKATPYTFPGGTGQILNKNRNLAEVWDE--FKNEFYIISPGVTK 390
 QY 332 IISPEYCMQV-----HICLPADIKLVKMSW 356
 DB 391 V-----DIQDISSRIGLRLKLCRCPPFSW 413

RESULT 7
 PACT_HUMAN
 ID PACT_HUMAN STANDARD: PRT: 559 AA.
 AC Q10472;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE POLYPEPTIDE N-ACETYLGLACTOSAMINYLTRANSFERASE (EC 2.4.1.41) (PROTEIN-
 DE UDP ACETYLGLACTOSAMINYLTRANSFERASE) (UDP-GALNAC:POLYPEPTIDE, N-
 DE ACETYLGLACTOSAMINYLTRANSFERASE) (GALNAC-T1).
 GN GALNT1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Salivary gland.
 RX MEDLINE=96115928; PubMed=8690719;
 RX Meurer J.A., Maylor J.M., Baker C.A., Thomsen D.R., Homa F.L.,

DB 129 GGEFTVGPPEFFILKEDACGSWVPADAG-----YFOLPELDGCDIRKIVFALENL 182
 QY 207 GCHT-----VAHIOHEVDHIFCMQV---IDVFOQKFGVEL----- 239
 DB 183 GHVASHHHEVAGGHEVDFKPADAVKTAADSVVTKTKITIAADGKATFMPKPFQI 242
 QY 240 -GESVAQIQAMMYKADPNIFTEREKRESAAYIPFGSDGYFIAALFGGTPT--OYLTNTQ 296
 DB 243 NQSGMCHQOSTIMLNCESPSY-----DESAQY-QLSETGMSVAGILIDHAKSIVAVTNPV 296
 QY 297 ECFKQILDKKNDIEQWDESHLNKYPFLNKP 329
 DB 297 NSYKRLVPGYEAFVNIAM--ANSNRSAIVRVP 326

RESULT 9

POL_SIVAI STANDARD; PRT: 1057 AA.
 AC 002836:
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-JUL-1993 (Rel. 26, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE POL. POLYPEPTIDE [CONTAINS: PROTEASE (RETROPEPTIN) (EC 3.4.23.16);
 REVERSE TRANSCRIPTASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)]
 GN POL.
 OS Simian immunodeficiency virus (isolate AGM / clone CRT-1) (SIV-AGM).
 OC Viruses; Retroviridae; Retroviridae; Lentivirus.
 OX NCBI_TaxID=31684:
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91220680; PubMed=2024476;
 RA Fomsgaard A., Hirsch V.M., Allan J.S., Johnson P.R.;
 RT A highly divergent proviral DNA clone of SIV from a distinct species
 of African green monkey.";
 RL Virology 182:397-402(1994).
 CC -1- PTM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE
 DETERMINED.
 CC -1- MISCELLANEOUS: THIS IS AN AFRICAN GREEN MONKEY ISOLATE.
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO
 KNOWN AS THE RETROPEPTIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DB EMBL: M66437; AAA91973.2;
 DB EMBL: M58410; CAB26041.1;
 DB HSSP: P03366; IHMV.
 DR MEROPS: A02.003;
 DR InterPro: IPR000477;
 DR InterPro: IPR001037;
 DR InterPro: IPR001584;
 DR InterPro: IPR001569;
 DR InterPro: IPR001995;
 DR InterPro: IPR002156;
 DR Pfam: PF00552; Integrase_Zn_1;
 DR Pfam: PF00075; rnselb_1;
 DR Pfam: PF00665; rve_1;
 DR Pfam: PF00077; rvp_1;
 DR Pfam: PF00078; rvt_1;
 DR PROSITE: PS00141; Asp_PROTEASE_1;
 DR PROSITE: PS50175; Asp_PROT_RETROV_1;
 KM AID5: Polyprotein; Hydrolyase; Aspartyl protease; Endonuclease;
 Nuclease; Transferase; RNA-directed DNA polymerase;
 FT ACT_SITE 114 114 BY SIMILARITY.
 SO SEQUENCE 1057 AA; 120566 MW; 0550E954011FAA2C CRC64;

Query Match 4.7%; Score 95; DB 1; Length 1057;

Best Local Similarity 22.6%; Pred. No. 6.5; Matches 72; Conservative 44; Mismatches 111; Indels 92; Gaps 19;

QY 69 INERKEQNESESKIKLSDMNPFRPEVYMTMKAPVWVCTYNAVLDNYAKK17 128
 DB 199 IEFKVKQLEKEDKPKLKW--PLSRKLEALTEICKQMEEGKLSIGENPV----- 250
 QY 129 VGLVFAVGYE---HYEEF--LTSANKHFV-----VGHPIVYIMVWVSMPLEL 178
 DB 251 -NTVFALKKKDKTQWMLVDFRRLNKATQDFPVQGLIPHPA---GLQKKQITVIDI 305
 QY 179 G-----PL--KSKFKYK-----PEKMQ-----DISMKMTIGBH 210
 DB 306 GDAYVSIPLCKEPRKYAFTIPSVNNTGPIRYQFNCLIQCKSPITFQNTAANILREI 365
 QY 211 VAH-----IOHEVDFEQMDVDQFQD-----KQVETLESVAQIQAM-W-- 250
 DB 366 KRHPGLEIVQYMDMLASDHDETRHQVDYRKMLLEKGLTPKKVGRREPPWEMG 425
 QY 251 YKADPNFTYERKRESAAYIPFGSDGYH--AIFGTPVOVNIITVECKGI----- 302
 DB 426 YKLHFNKMTINKIE-----LPLHSEFWTVNKLQKVG-----VINMASQIYPGIKTKHTC 475
 QY 303 --LKKRNDI-EAQWDESH 318
 DB 476 AMLRCKKNMLDEPIWTFEA 494

RESULT 10

PAGT_RAT STANDARD; PRT: 559 AA.

AC Q10473:
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE POLYPEPTIDE N-ACETYLGLYCOSAMINYLTRANSFERASE (EC 2.4.1.41) (PROTEIN-
 UDP ACETYLGLYCOSAMINYLTRANSFERASE) (UDP-GALNAc:POLYPEPTIDE, N-
 ACETYLGLYCOSAMINYLTRANSFERASE) (GALNAC-TT).
 GN GALNT1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116:
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Sublingual gland;
 RX MEDLINE=96318029; PubMed=8748168;
 RA Hagen F., Gregorie C.A., Tabak L.A.;
 RT *Cloning and sequence homology of a rat UDP-GalNAc:polypeptide N-
 acetylglucosaminyltransferase.";
 RL Glycoconj. J. 12:901-909(1995).
 CC -1- FUNCTION: THIS PROTEIN CATALYZES THE INITIAL REACTION IN O-LINKED
 OLIGOSACCHARIDE BIOSYNTHESIS, THE TRANSFER OF AN N-ACETYL-D-
 GALACTOSAMINE RESIDUE TO A SERINE OR THREONINE RESIDUE ON THE
 PROTEIN RECEPTOR.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GALACTOSAMINE + POLYPEPTIDE =
 UDP + N-ACETYL-D-GALACTOSAMINYL-POLYPEPTIDE.
 CC -1- COFACTOR: MANGANESE AND CALCIUM.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN, Golgi.
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, SPLEEN, LIVER, SKELETAL MUSCLE
 AND KIDNEY.
 CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

RL FASEB J. 5:2052-2060(1991).
 RN [111]
 RP REVIEW ON VARIANTS.
 RX MEDLINE-97255959; PubMed 9101299;
 RA Kivimäki H., Timp G., Prockop D.J.:
 RT "Mutations in fibrillar collagen (types I, II, III, and XI), fibrillar
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels."
 RL Hum. Mutat. 9:300-315(1997).
 RN [112]
 RP VARIANT SER-1074.
 RX MEDLINE-90046909; PubMed 2572591;
 RA Vissing H., O'Aleese M., Lee H., Ramirez F., Godfrey M.,
 RA Hollister D.W.:
 RT "Glycine to serine substitution in the triple helical domain of pro-
 RT alpha 1 (I) collagen results in a lethal perinatal form of short-
 RT limbed dwarfism."
 RL J. Biol. Chem. 264:18265-18267(1989).
 RN [113]
 RP VARIANT SER-1095--TYR-1440 DEL.
 RX MEDLINE-89266907; PubMed 2544071;
 RA Lee H., Vissing H., Ramirez F., Rogers D., Knecht D.:
 RT "Identification of the molecular defect in a family with
 RT spandyloepiphyseal dysplasia."
 RL Science 244:978-980(1989).
 RN [114]
 RP VARIANT OSTEOARTHRITIS CYS-650.
 RX MEDLINE-90370826; PubMed 1975694;
 RA Ala-Kokko L., Baldwin G.T., Moskowitz R.W., Prockop D.J.:
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a
 RT cause of primary osteoarthritis associated with a mild
 RT chondrodysplasia."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:6505-6508(1990).
 RN [115]
 RP VARIANT GLY-1911-36.
 RX MEDLINE-94291136; PubMed 2064612;
 RA Bateman J.F., Hamman M., Chan D., Cole W.G.:
 RT "Characterization of a type I collagen alpha 2(I) glycyne-586 to
 RT valine substitution in osteogenesis imperfecta type IV. Detection of
 RT the mutation and prenatal diagnosis by a chemical cleavage method."
 RL Biochem. J. 276:765-770(1991).
 RN [116]
 RP VARIANT OSTEOARTHRITIS CYS-650.
 RX MEDLINE-91086471; PubMed 1985108;
 RA Eyre D.R., Weiss M.A., Moskowitz R.W.:
 RT "Cartilage expression of a type II collagen mutation in an inherited
 RT form of osteoarthritis associated with a mild chondrodysplasia."
 RL J. Clin. Invest. 87:457-461(1991).
 RN [117]
 RP VARIANT HYPERCHONDROGENESIS G10-984.
 RX MEDLINE-94054749; PubMed 1429602;
 RA Bouquet R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.:
 RT "An amino acid substitution (G1853-->Glu) in the collagen alpha
 RT 1(I) chain produces hyperchondrogenesis."
 RL J. Biol. Chem. 267:22522-22526(1992).
 RN [118]
 RP VARIANT HYPERCHONDROGENESIS SER-705.
 RX MEDLINE-92262484; PubMed 147406;
 RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
 RA Ramirez F., Vitale E., Lee H.:
 RT "Characterization of a type II collagen gene (COL2A1) mutation
 RT identified in cultured chondrocytes from human hyperchondrogenesis."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
 RN [119]
 RP VARIANT WS-11 ASP-198.
 RX MEDLINE-93404428; PubMed 8417498;
 RA Kivimäki H., Kivimäki P., Härdala L., Kivimäki H.,
 RA Kivimäki K.L., Prockop D.J., Ala-Kokko L.:
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
 RT for glycine alpha 1-67 and that causes cataracts and retinal
 RT detachment: evidence for molecular heterogeneity in the Wagner
 RT syndrome and the Stickler syndrome (arthro-ophthalmopathy)."

RL Am. J. Hum. Genet. 53:55-61(1993).
 RN [120]
 RP VARIANT SER-840.
 RA Tiller G.E., Weiss M.A., Lachman R.S., Cohen D.E., Rimoin D.L.:
 RA Eyre D.R.:
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces
 RT spandyloepiphyseal dysplasia (SEMP), Strudwick type."
 RL Am. J. Hum. Genet. 53:A209-A209(1993).
 RN [121]
 RP VARIANT OSTEOARTHRITIS CYS-650.
 RX MEDLINE-93282819; PubMed-8507190;
 RA Holdersum D., Malmend G.J., Moskowitz R.W., Baggi T.M.:
 RT "Human cartilage from late stage lamellar osteoarthritis transcripts
 RT type II collagen mRNA encoding a cysteine in position 519."
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
 RN [122]
 RP VARIANT SER-285.
 RX MEDLINE-93252400; PubMed-8486475;
 RA Viskochil M., Kivimäki P., Vuorio A.F., Kivimäki L., Ala-Kokko L.:
 RA Polonon L.:
 RT "A mutation in the amino-terminal end of the triple helix of type II
 RT collagen causing severe osteochondrodysplasia."
 RL Genomics 16:282-285(1993).
 RN [123]
 RP VARIANT SER-206.
 RX MEDLINE-94063862; PubMed-8244341;
 RA Williams G.J., Considine E.L., Knowlton R.G., Reigato A., Neumann G.,
 RA Harcourt D., Buxton P., Jimenez S.A., Prockop D.J.:
 RT "Spandyloepiphyseal dysplasia and precocious osteoarthritis in a
 RT family with an Arg75-->Cys mutation in the procollagen type II gene
 RT (COL2A1)."
 RL Hum. Genet. 92:499-505(1993).
 RN [124]
 RP VARIANT SER-920.
 RX MEDLINE-93315508; PubMed-8325895;
 RA Chan D., Taylor T.K.F., Cole W.G.:
 RT "Characterization of an arginine 789 to cysteine substitution in
 RT alpha 1 (I) collagen chains of a patient with spandyloepiphyseal
 RT dysplasia."
 RL J. Biol. Chem. 268:15238-15245(1993).
 RN [125]
 RP VARIANT SER-1128.
 RX MEDLINE-93140139; PubMed-8423604;
 RA Cole W.G., Ball R.K., Rogers J.G.:
 RT "The clinical features of spandyloepiphyseal dysplasia congenita
 RT resulting from the substitution of glycine 947 by serine in the alpha
 RT 1(I) chain of type II collagen."
 RL J. Med. Genet. 30:27-35(1993).

Query Match 4.68; Score 92.5; DH 1; Length 1418;
 Best local similarity 18.78; Pred. No. 15;
 Matches 45; Conservative 25; Mismatches 50; Indels 121; Gaps 10;

QY 23 EYHSPESL-----FWINPSR-----NPEVSSSI--48
 DB 1198 ESTSPESKRNPATCKDLKICHPKMSQYWDIPNCLTIDAMKYVCNMECTCYVP 1257
 QY 49 -----OKGWLPR-----WF-----NNGYEDKQDINEKKEQNEHESKILKSDWNP 91
 DB 1258 NPANPKKNWSSKSKKKHIFGELINGGFHSYGDNLAPNTANVQMTFLRL----- 1311
 QY 92 FKRPVIVMTKKAIVVWGGYTNRAVLNYYAKOKLIVGLIVFAVGRVIEHYIEFTLSA 151
 DB 1312 -----LSTEGSON-----ITFKKNSIA-----YLDFAACNL 1338
 QY 152 NKRENV-----GHVPLVTVMDV-----VSPKQVLELCP 180
 DB 1339 KALLIGSNOVELIACGNSRFTYALKIKGCKIKHGKMKQVIVYRSQKTSRLDIP 1398
 QY 181 L 181
 DB 1399 M 1399

RESULT 14
ID HEXL_ENTHI STANDARD; PRT; 522 AA.
AC P49009;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE BETA-HEXOSAMINIDASE (EC 3.2.1.52) (N-ACETYL-BETA-D-GLUCOSAMINIDASE)
DE (BETA-GLUCANASE) (BETA-N-ACETYLHEXOSAMINIDASE).
GN HEX-A1 OR HEX-1.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HM-11MSS;
RX MEDLINE=96010595; PubMed=7581340;
RA Beaman M.J., Bailey G.B.;
RT *The primary structure of an Entamoeba histolytica
RT beta-hexosaminidase A subunit.*;
RL J. Eukaryot. Microbiol. 42:632-636(1995).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING N-ACETYL-
CC D-HEXOSAMINE RESIDUES IN N-ACETYL-BETA-D-HEXOSAMINIDES.
CC -1- SIMILARITY: BELONGS TO FAMILY 20 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U09735; AAA80165.1; -
DR HSSP: P06865; 10BC.
DR InterPro: IPR001540; -
DR Pfam: PF00728; Glyco_hydro.20.1.
DR PRINTS: PR00738; GLYHYDRLASE20.
DR Hydrolase: Glycosidase.
KM SEQUENCE 522 AA; 59436 MW; B793C84B565A83AB CRC64;
SU
Query Match 4.5%; Score 90; DB 1; Length 522;
Best Local Similarity 20.5%; Pred. No. 6.7;
Matches 55; Conservative 38; Mismatches 101; Indels 74; Gaps 14;
QY 17 VIVFWEYIHSPE---GSLFWINPSRN-----EYGGSSIOK 50
DB 246 VVANCMDIYVTSMRGCVNLINPA-NPRTFPIIALMKELSDFTGTYVHAGDEV--- 302
QY 51 GWWLPPWPNNGYHREDDINERKQPNDESKLSDWTFNPKRPVYVMTKKAVYME 110
DB 303 -WTSQMSK---KEVSDIQKFKMSKGLN-SLLEEGYRNKVAQOEVI--HMKHHPVME 354
QY 111 GTNRAVLDN-----YAKQKITVGL-TVFVAGRTIENYLE-----EFLTSANKH 154
DB 355 EYFKKGNADKNITIIQVWDDIRLQGVVNSQYKAIFSGAFYLDKQMLCNSYSSTCVNTH 414
QY 155 FWGGRVITFVIVDDVSRMPLIELGLRSFKVF-----KIKPKRQDISMMMKIT 206
DB 415 SM-----WWTNEDMTDNDPVSFKLSSEKENVLGGEGCSNGESTDEQNFEDRVYKISAI 469
QY 207 GEH-----IVAHIOHEV--DPLFCMDV 226
DB 470 AERLMSKESVVDKESHEVRANTRLCIDV 497
RESULT 15
ID TALA_SV40 STANDARD; PRT; 708 AA.
AC F03070;
DT 21-JUL-1986 (Rel. 01, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE LARGE T ANTIGEN.
OS Simian virus 40 (SV40).
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Polyomavirus.
OX NCBI_TaxID=10633;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=78159686; PubMed=205947;
RA Reddy V.B., Thimmappa R., Dhar R., Subramanian K.N., Zain R.S.,
RA Pan J., Ghosh P.K., Ceima M.L., Weissman S.M.;
RT *The genome of simian virus 40.*;
RL Science 200:494-502(1978).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=776.
RX MEDLINE=78156432; PubMed=205802;
RA Fiers W., Contreas R., Haegeman G., Rogiers R., van de Voorde A.,
RA van Heuverswyn H., van Herrewege J., Volckaert G., Ysebaert M.;
RT *Complete nucleotide sequence of SV40 DNA.*;
RL Nature 273:113-120(1978).
RN [3]
RP NUCLEAR LOCALIZATION SIGNAL.
RX MEDLINE=85074458; PubMed=6096007;
RA Kalderson D., Roberts B.L., Richardson W.D., Smith A.F.;
RT *A short amino acid sequence able to specify nuclear location.*;
RL Cell 39:499-509(1984).
RN [4]
RP ZINC-FINGER DOMAIN.
RX MEDLINE=91056604; PubMed=2173794;
RA Hoess A., Moorel I.F., Fanning E., Arthur A.K.;
RT *The finger domain of simian virus 40 large T antigen controls
RT DNA-binding specificity.*;
RL J. Virol. 64:6291-6296(1990).
RN [5]
RP ZINC-FINGER DOMAIN.
RX MEDLINE=91237835; PubMed=1851875;
RA Loebner G., Stenger J.E., Ray S., Parsons R.F., Anderson M.E.,
RA Tegetmeyer P.;
RT *The zinc finger region of simian virus 40 large T antigen is needed
RT for hexamer assembly and origin melting.*;
RL J. Virol. 65:3167-3174(1991).
RN [6]
RP ASSOCIATION WITH HOST RB PROTEIN.
RX MEDLINE=88270506; PubMed=2839300;
RA Decaprio J.A., Ludlow J.W., Figue J., Shew J.-Y., Huang C.-M.,
RA Lee W.-H., Marsilio E., Paucha E., Livingston D.M.;
RT *SV40 large tumor antigen forms a specific complex with the product of
RT the retinoblastoma susceptibility gene.*;
RL Cell 54:275-283(1988).
RN [7]
RP IN VITRO PHOSPHORYLATION.
RX MEDLINE=88265851; PubMed=2838952;
RA Graesser F.A., Scheidmann K.H., Tuazon P.T., Traugh J.A., Walter G.;
RT *In vitro phosphorylation of SV40 large T antigen.*;
RL Virology 165:13-22(1988).
RN [8]
RP SEQUENCE OF 102-118, AND PHOSPHORYLATION SITE THR-124 BY CDC2.
RX MEDLINE=90015163; PubMed=2552322;
RA McVey D., Britzuela L., Mohr I., Marshak D.R., Gluzman Y., Beach D.;
RT *Phosphorylation of large tumour antigen by cdc2 stimulates SV40 DNA
RT replication.*;
RL Nature 341:503-507(1989).
RN [9]
RP IN VITRO ACTIVATION OF DNA REPLICATION BY PP2A.
RX MEDLINE=90059993; PubMed=255176;
RA Virshup D.M., Kaufman M.G., Kelly T.J.;
RT *Activation of SV40 DNA replication in vitro by cellular protein
RT phosphatase 2A.*;
RL EMBO J. 8:3891-3898(1989).
RN [10]
RP REGULATION OF REPLICATION BY PHOSPHORYLATION.
RX MEDLINE=90263097; PubMed=2160857;

RA. p105. C-1.
 RT. "The replication functions of SV40 T antigen are regulated by phosphorylation."
 RL. Cell 61:745-758(1990).
 RN. (11).
 RP. DEPHOSPHORYLATION BY HOST PP2A.
 RX. MEDLINE-91172166; PubMed 1849668;
 RA. Scholtenmann K.H., Mumby M.C., Kende-H K., Walter G.;
 RT. "Dephosphorylation of simian virus 40 large-T antigen and p53 protein by protein phosphatase 2A: inhibition by small-t antigen."
 RL. Mol. Cell. Biol. 11:1996-2004(1991).
 RN. (12).
 RP. SITES OF DEPHOSPHORYLATION BY HOST PP2A.
 RX. MEDLINE-91162764; PubMed 1848420;
 RA. Scholtenmann K.H., Virshup D.M., Kelly T.J.;
 RT. "Protein phosphatase 2A dephosphorylates simian virus 40 large T antigen specifically at residues involved in regulation of DNA-binding activity."
 RL. J. Virol. 65:2098-2101(1991).
 RN. (13).
 RP. ACTIVATION OF REPLICATION BY PHOSPHORYLATION ON THR-124.
 RX. MEDLINE-96211524; PubMed 8648725;
 RA. McVey D., Winkler B., Tegtmeyer P.;
 RT. "Mechanisms of simian virus 40 T-antigen activation by phosphorylation of threonine 124."
 RL. J. Virol. 70:3887-3893(1996).
 CC. -1- FUNCTION: INITIATES DNA UNWINDING AND REPLICATION VIA ELABORATE INTERACTIONS WITH THE VIRAL ORIGIN OF REPLICATION. BINDS TWO ADJACENT SITES IN THE SV40 ORIGIN.
 CC. -1- SUBUNIT: FORMS HOMODIMERS IN THE PRESENCE OF ATP.
 CC. -1- SUBCELLULAR LOCATION: NUCLEAR.
 CC. -1- DOMAIN: THE ZINC FINGER REGION CONTRIBUTES TO PROTEIN-PROTEIN INTERACTIONS ESSENTIAL FOR THE ASSEMBLY OF STABLE T-ANTIGEN HEXAMERS AT THE ORIGIN OF REPLICATION AND HEXAMERS ARE NEEDED FOR SUBSEQUENT ALTERATIONS IN THE STRUCTURE OF ORIGIN DNA.
 CC. -1- PTM: PHOSPHORYLATED ON BOTH SERINE AND THREONINE RESIDUES. PHOSPHORYLATION ON SER-120 AND SER-124 INHIBITS VIRAL REPLICATION, WHILE THAT ON THREONINE-124 ENHANCES REPLICATION BY ACTIVATING THE DNA BINDING DOMAIN.
 CC. -1- PTM: DEPHOSPHORYLATED PREFERENTIALLY BY PP2A ON SERINE RESIDUES 124, 125, 677 AND PERHAPS 679. SMALL T ANTIGEN INHIBITS THE DEPHOSPHORYLATION BY THE AC FORM OF PP2A.
 CC. -1- MISCELLANEOUS: THE INITIAL 82 RESIDUES OF THE LARGE T AND SMALL T ANTIGENS FROM SV40 ARE COMED BY THE SAME NOCTURNE SEQUENCE.
 CC. This SWISS Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb.stb.ch/announce/> or send an email to licensing@isb.stb.ch).
 DR. EMBL: V01480; CAA24672.1;
 DR. EMBL: J02400; AAB59924.1;
 DR. F1R: A04607; TWP474.
 DR. TRANSFAC: T00788;
 DR. Glycerol: P04070;
 DR. InterPro: IPR001623;
 DR. Pfam: PF00226; Pfam1.1;
 DR. PROSITE: PS00076; DNAS2.1;
 KW. Early protein; Acetylation; Nuclear protein; phosphorylation;
 KM. ATP binding; DNA-binding; DNA replication; Zinc-finger;
 FT. MOD_RES 1 114
 FT. DOMAIN 105 114
 FT. INMAIN 125 142
 FT. DNA_BIND 131 259
 FT. DOMAIN 286 289
 FT. ZN_FING 302 320
 FT. DOMAIN 347 672
 FT. INMAIN 426 433
 FT. NIT_BIND 433 638
 FT. INMAIN 633 638
 POLY-ASP.
 POLY-LEU.
 C2H2-TYPE (ATYPICAL).
 BINDING TO HOST P53 PROTEIN.
 ATP (POTENTIAL).
 POLY-ASP.

FT	MOD_RES	106	106	PHOSPHORYLATION.
FT	MOD_RES	112	112	PHOSPHORYLATION.
FT	MOD_RES	120	120	PHOSPHORYLATION.
FT	MOD_RES	123	123	PHOSPHORYLATION.
FT	MOD_RES	124	124	PHOSPHORYLATION.
FT	MOD_RES	639	639	PHOSPHORYLATION.
FT	MOD_RES	676	676	PHOSPHORYLATION.
FT	MOD_RES	677	677	PHOSPHORYLATION.
FT	MOD_RES	679	679	PHOSPHORYLATION.
FT	MOD_RES	701	701	PHOSPHORYLATION.
FT	MOD_RES	531	531	F -> Y (IN STRAIN 776).
FT	VARIANT	549	549	A -> P (IN STRAIN 776).
FT	VARIANT	552	552	Y -> P (IN STRAIN 776).
FT	MUTAGEN	124	124	T -> A: 200 FOLD REDUCTION IN PHOSPHORYLATION BY CACZ. NO DNA REPLICATION ACTIVATION.
FT	MUTAGEN	679	679	S -> A: ENHANCED DNA REPLICATION IN VIVO.
FT	SEQUENCE	708 AA:	81582 MW:	CHB1306F9EAEZC0 C8C64;

Query Match Best local similarity 4.78; Score 90; DB 1; Length 708;
 Matches 59; Conservative 54; Mismatches 118; Indels 110; Gaps 15;

QY	23	RYHSPPKSLFWINPNSRNPVEYSSISOKWMLPRWFNNNGYHEHGDINEKPDHNEPSK	82
DB	67	KYAGPPOGGEW-DATETPTYGDEMD-WM-----NATNEMLPESSEKPS-SIDPAT	117
QY	83	LKLSDFEPFRPREPVMTKKKAPVWEGFYNAVILN-----YAKQKITGLTIV	134
DB	118	ADSOHSTPPKKKKRKEVPKDPSEL--SPLSHAVFSKTLACPAIVTKRK-----AAL	170
QY	135	AVGYIEHTLEELTSANKHDMVGLPVF-----	163
DB	171	LYKKIMKESVTFI---SRHNSYNNITLFLTPKIRVSAINNTAQKICTPSFLICGVN	227
QY	164	--YIMDVDSMP--LIEGPIRSKPVKIKER-----RMQDISMRKKTIGEHIV	211
DB	228	KEYLVSALTRDPSSVIEESLPGKLKEIDNPENEKEETKQVSKLVTEYAFKCDIVL	287
QY	212	---AUIHEVDFLPGMDVQVPOKESVETLGSVAJQAWMYKADINQTYERRKESAA	268
DB	288	LIGVTLRFQYSEMKIKIC-----KQPSHYKVRK-----	319
QY	269	YTPGEGPYHAALFGSTPTGVNITDGECKGLIKDKND	309
DB	320	-----HYAAALFADSKND-KTICQADVTLAKKIV	351

Search completed: November 5, 2001, 15:24:46
 Job time: 3258 sec